Statistical analysis of patient-reported outcome data in randomised controlled trials of locally advanced and metastatic breast cancer

Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data Consortium (SISAQOL); Pe, Madeline; Calvert, Melanie

DOI:
10.1016/S1470-2045(18)30418-2

Document Version
Peer reviewed version

Citation for published version (Harvard):
Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data Consortium (SISAQOL), Pe, M & Calvert, M 2018, 'Statistical analysis of patient-reported outcome data in randomised controlled trials of locally advanced and metastatic breast cancer: a systematic review', *The Lancet Oncology*, vol. 19, no. 9, pp. e459-e469. https://doi.org/10.1016/S1470-2045(18)30418-2

Current State of Statistical Analysis of Patient Reported Outcomes Data in Cancer Randomized Controlled Trials on Locally Advanced and Metastatic Breast Cancer – A Systematic Review

Madeline Pe[1], Lien Dorme[1], Corneel Coens[1], Ethan Basch[2], Melanie Calvert[3], Alicyn Campbell[4], Charles Cleeland[5], Kim Cocks[6], Laurence Collette[1], Linda Dirven[7], Amylou C Dueck[8], Nancy Devlin[9], Hans-Henning Flechtner[10], Carolyn Gotay[11], Ingolf Griebsch[12], Mogens Groenvold[13], Madeleine King[14], Michael Koller[15], Daniel C Malone[16], Francesca Martinelli[1], Sandra A Mitchell[17], Jammbe Z Musoro[1], Kathy Oliver[18], Elisabeth Piault-Louis[4], Martine Piccart[19], Francisco L Pimentel[20], Chantal Quinten[21], Jaap C Reijneveld[22], Jeff Sloan[23], Galina Velikova[24], and Andrew Bottomley[1]

for the Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data (SISAQOL) Consortium.

[1]European Organisation for Research and Treatment of Cancer (EORTC), Brussels, Belgium

[2]Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA

[3]Centre for Patient Reported Outcomes Research, Institute of Applied Health Research, College of Medical and Dental Sciences, University of Birmingham, UK

[4]Genentech, a member of the Roche group, San Francisco, CA, USA

[5]Department of Symptom Research, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

[6]Adelphi Values, Bollington, Cheshire, UK

[7]Leiden University Medical Center/Haaglanden Medical Center, Leiden/The Hague, Netherlands

[8]Alliance Statistics and Data Center, Mayo Clinic, Scottsdale, AZ, USA

[9]Office of Health Economics, London, UK

[10]Clinic for Child and Adolescent Psychiatry and Psychotherapy, University of Magdeburg, Magdeburg, Germany

[11]School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada

[12]Boehringer Ingelheim International GmBH, Ingelheim, Germany

[13]Department of Public Health and Bispebjerg Hospital, University of Copenhagen, Copenhagen, Denmark

[14]School of Psychology and Sydney Medical School, University of Sydney, Sydney, NSW, Australia

[15]Center for Clinical Studies, University Hospital Regensburg, Regensburg, Germany

[16]College of Pharmacy, University of Arizona, Tucson, AZ, USA

[17]Outcomes Research Branch, Healthcare Delivery Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD, USA

[18]International Brain Tumour Alliance, Surrey, UK

[19]Institut Jules Bordet, Université Libre de Bruxelles (ULB), Brussels, Belgium

[20]Blueclinical Phase I, Porto, Portugal; Centro de Estudos e Investigação em Saúde da Universidade de Coimbra, Portugal.

[21]European Centre for Disease Prevention and Control, Surveillance and Response Support Unit, Epidemiological Methods Section, Stockholm, Sweden

[22]VU University Medical Center, Department of Neurology & Brain Tumor Center, Amsterdam, The Netherlands

[23]Alliance Statistics and Data Center, Mayo Clinic, Rochester, MN, USA

[24]Leeds Institute of Cancer and Pathology, University of Leeds, St James's Hospital, Leeds, UK.

Corresponding author: Dr. Madeline Pe

Quality of Life Department, European Organisation for Research and Treatment of Cancer

Avenue E. Mounier 83/11, 1200 Brussels, Belgium

madeline.pe@eortc.org

Tel: +32 2 774 1544

Keywords

Health-Related Quality of Life, Advanced Breast Cancer, Systematic Review, Randomized Controlled Trials, Statistical Methodology, Patient Reported Outcomes

Summary

Although patient reported outcomes (PROs) such as health-related quality of life (HRQOL) are important endpoints in randomized controlled trials (RCTs), there is little consensus about analysis, interpretation and reporting of these data.

A systematic review was conducted to assess variability, quality, and standards of PRO data analyses in advanced breast cancer RCTs. We searched through PubMed for English language articles published in peer-reviewed journals between January 2001 and October 2017. Eligible articles reported PRO results from RCTs involving adult advanced breast cancer patients receiving anti-cancer treatments with reported sample sizes of at least 50 patients.

Sixty-six RCTs met the selection criteria. A small number of RCTs reported a specific PRO research hypothesis (8/66, 12%). There was heterogeneity in the statistical methods used to assess PRO data, with a mixture of longitudinal and cross-sectional techniques. Not all articles addressed the problem of inflated type I error resulting from multiple testing. Fewer than half of RCTs reported the clinical significance of their findings (28/66, 42%). The majority of trials did not report how missing data was handled (48/66, 73%).

Our review demonstrates a need to improve standards in analysis, interpretation and reporting of PRO data in cancer RCTs. Lack of standardization makes it difficult to draw robust conclusions and compare findings across trials. The Setting International Standards in the Analyzing Patient-Reported Outcomes and Quality of Life Data (SISAQOL) Consortium was set up to address this need and develop recommendations on the analysis of PRO data in RCTs.

Acknowledgments

We would like to acknowledge Irina Ghislain for her help in the initial extraction of the relevant papers. We would also like to acknowledge Dr. Mariana Brandao and Dr. Noam Pondé for their help in extracting the treatment arms and pointing out which trials are practice changing.

We would like to dedicate this manuscript to SISAQOL consortium member, Franck Bonnetain, who passed away in May 2017.

Financial support for this research was provided by the EORTC Cancer Research Fund and by an unrestricted education grant from Boehringer Ingelheim.

Introduction

In a breakthrough report, the Institute of Medicine highlighted *patient-centered care* as a critical component of quality health care[1]. Patient-centered care is defined as "respectful of, and responsive to the individual patient preferences, needs, and values and that patient values guide all clinical decisions" [1]. The incorporation of patient reported outcomes (PROs) in randomized controlled trials (RCTs) is one concrete way of responding to this imperative. Increasingly, PRO endpoints are being included in RCTs to assess clinical benefit alongside overall and progression-free survival[2]. PRO is any outcome that is reported directly by the patient[3,4]. By including PRO endpoints, such as health-related quality of life (HRQOL), the patient's perspective is obtained, providing better patient information and supporting shared decision making in the development of new therapies[5,6].

However, the lack of standards and clear guidelines on how these patient-reported data should be analyzed and interpreted in RCTs diminishes their recognized and important value by making it difficult to compare results across trials and draw conclusions about the patient experience of new types of cancer treatment[7]. Data generated from certain PROs, such as HRQOL, are complex: they (a) are multidimensional, with several subscales to characterize patients' symptoms and their impact on aspects of patient functioning; (b) require repeated measurements in order to capture changes in these outcomes; and (c) are prone to missing data since it is often difficult to obtain complete PRO follow-up data from all randomized patients[8,9]. Inappropriate handling of these critical statistical issues could bias findings and lead to inaccurate conclusions. Current guidelines do not provide concrete suggestions on how to deal with statistical issues concerning PROs and need to be supplemented with more detailed strategies on how to address these concerns[3,10].

The Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data for Cancer Clinical Trials (SISAQOL) Consortium was established to respond to a clear need to develop standards, guidelines, and recommendations for the analyses of PRO data in cancer RCTs. This Consortium involves a wide range of international experts - leading PRO researchers and statisticians as well as key individuals from different international oncological and medical societies, advisory and regulatory bodies, academic societies, the pharmaceutical industry, cancer institutes, and patient advocacy organizations[11]. A key task identified by the Consortium was to undertake systematic literature reviews to describe the current state of PRO analyses in RCTs of cancer treatment. The current article examines how analyses of PRO such as HRQOL are conducted in RCTs, in this case using anti-cancer treatments for advanced breast cancer as an example set of trials commonly seen in the literature. Since maintaining HRQOL is important in the care of advanced breast cancer patients, it was a reasonable expectation that a considerable number of advanced breast cancer RCTs would have included PROs in their assessments [12].

Methods

Search strategy and selection criteria

We followed the methodology noted in the guidelines for the Cochrane Handbook for Systematic Reviews of Interventions[13] and the results of this review are reported in accordance with PRISMA guidelines (see Appendix page 35-36 for the PRISMA checklist) [14]. We did not publish a review protocol for this study. A literature search was performed in PubMed on March 30, 2016 (and updated on February 7, 2018) with the following keywords: (quality of life[MeSH Terms] OR quality of life[Text Word] OR patient reported outcomes[Text Word]) AND (advanced[All Fields] OR metastatic[All Fields]) AND breast cancer[Text Word] AND (Randomized Controlled Trial) AND (breast neoplasm[MeSH Terms]) AND (Clinical Trial[ptyp] AND ("2001/01/01"[PDat] : "2017/10/30"[PDat]) AND Humans[Mesh]). Using this search strategy, 323 potentially eligible articles were identified. Checking of references of publications were also undertaken. In addition, we performed a Web of Science search at a later date (April 22, 2018), but no further articles were found.

The inclusion and exclusion criteria for the RCTs were similar to that of Ghislain and colleagues[15]. The inclusion criteria were: articles should report PRO findings from RCTs involving adult advanced breast cancer patients (18 years or older), receiving anti-cancer treatments (chemotherapy, targeted therapy, endocrine therapy) with sample sizes of at least 50 patients. Advanced breast cancer refers to either metastatic breast cancer or locally advanced breast cancer (see ESO-ESMO international consensus guidelines for more information)[12]. Only articles published in a peer-reviewed journal between January 2001 and October 2017 were included, regardless of starting or completion date of the study. It was originally considered to do

a search from 1997 to have exactly 20 years of review. However, due to the difficulty of retrieving articles before 2001, it was decided to begin the search from 2001.

Exclusion criteria were any RCTs which evaluated psychological, supportive or supplementary interventions. Supplementary treatments were defined as any other interventions that did not include anti-cancer therapy. Purely methodological or review publications were also excluded. Quality-adjusted life years (QALY) endpoints were not considered as PRO endpoints. Publications that reported interim analysis or the analyses of subgroups of patients (i.e., subgroups within the PRO cohort) were excluded since we wanted to limit the reporting to the top-level PRO results of the RCTs. Figure 1 presents the search strategy flowchart and the inclusion and exclusion criteria.

Two reviewers (MP and LDo) received the initial list of the 323 potentially eligible articles and the list of inclusion and exclusion criteria. They independently screened the articles based on these criteria. One reviewer (LDo) checked both assessments for any disagreements. Any disagreements were resolved through discussion. A third reviewer (CC) was available when no consensus could be reached.

[Insert Figure 1 here]

Evaluation criteria were adapted from previous reviews[16,17] with adjustments to enable in-depth assessment of statistical issues critical for PRO analysis. The initial data extraction sheet was developed by MP and CC and pilot-tested on three randomly-selected included studies and was further refined. This resulted in 23 evaluation criteria, classified into five broad categories: (1)

general description of the article, (2) reporting of research objectives, (3) statistical analysis and clinical relevance, (4) baseline assessment, and (5) assessing the amount of, and handling of missing data (see Appendix, page 29-34, for more details on the list of variables that were extracted). Two reviewers (MP and LDo) independently evaluated all identified studies on this predefined checklist of 23 criteria. One reviewer (LDo) checked the completed data extraction sheets for any disagreements. In case of disagreement, the article was reassessed by both reviewers together. If no consensus could be reached, a third reviewer (CC) served as a mediator to resolve disagreements.

When multiple publications for one RCT were identified, the article with the more comprehensive PRO statistical reporting was included in the review (see articles with bold formatting in the Appendix, page 1-28).Therefore, findings reported in this systematic review are based on the number of unique RCTs.

Results

Table 1 summarizes the overall main findings of this systematic review. To assess whether practices were improving over time, results were grouped into three periods (2001-2006; 2007-2012; 2013-2017) in Table 2. Details about individual papers included in this review are in the Appendix, page 1-28.

Descriptive Statistics

The search identified 335 eligible articles, of which a total of 66 eligible RCTs in advanced breast cancer were included, involving a total of 26,905 patients. No disagreements occurred between the 2 independent reviewers. The sample size ranged between 66 and 1102, with an average of 407. From the 66 trials, 12 were considered to be practice changing trials. The most commonly used PRO measures were two cancer-specific HRQOL questionnaires: the EORTC QLQ-C30 (35/66, 53%) and the FACT-B (22/66, 33%). Almost half of the RCTs (27/66, 41%) used multiple assessment tools to measure PROs, of which six trials (6/27, 22%) used an instrument that was not validated (e.g., ad-hoc trial specific checklists) in addition to a validated questionnaire. The majority of the PRO endpoints were reported as secondary endpoints (46/66 trials; 70%), with only three RCTs using a PRO as a primary endpoint (3/66, 5%). The other RCTs either reported PRO as an exploratory endpoint (3/66, 5%) or did not clearly report the PRO endpoint (14/66, 21%).

[Insert Table 1 here]

Reporting of research objectives

Only eight of 66 RCTs (12%) reported a hypothesis specific enough to inform the analysis of the PRO endpoint (i.e., the direction of hypothesis is stated with the domain of interest and specified time frame). The majority of the articles either reported a broad hypothesis (25/66, 38%; e.g., "to evaluate HRQOL between treatment arms") or no hypothesis (33/66, 50%). The majority of RCTs failed to report a specific PRO hypothesis, and there was no consistent improvement over time (2001-2006: 0/20, 0%; 2007-2012: 4/24, 17%; 2013-2017: 4/22, 18%).

Statistical analysis and clinical relevance

The majority of the trials (59/66, 89%) reported analyzing multivariate data, with multiple PRO scales/domains and/or with repeated assessments, to assess the PRO endpoint. Scales/domains refer to PRO variables that were analyzed in the trial. Thirty-eight RCTs analyzed multiple PRO scales/domains (38/66, 58%); and 21 RCTs analyzed a single PRO scale/domain (21/66, 32%). Among the 38 RCTs that used multiple PRO scales/domains, only six employed a statistical correction to correct for multiple testing (6/38, 16%). Two RCTs reported PROs as an exploratory endpoint and assessed multiple outcomes. It can be argued that exploratory endpoints do not have to correct for multiple testing. Results remained relatively the same after removing these two exploratory endpoints from the total score of PROs that assessed multiple outcomes (6/36, 17%). Combined, these numbers demonstrate that 27 of the 66 trials (41%) addressed the issue of multiple testing either by statistically correcting for multiple scales/domains or assessing only one scale/domain (often identified *a priori* as the most relevant scale/domain). There was no clear pattern in these findings (2001-2006: 11/20, 55%; 2007-2012: 7/24, 29%; 2013-2017: 9/22, 41%).

Fifty-three RCTs analyzed data with repeated assessments at follow-up (>1 follow-up assessment; 53/66, 80%); and 8 RCTs analyzed data with a single follow-up assessment (8/66, 12%). Among the RCTs that used multiple follow-up assessment points in their primary PRO analysis, 33 RCTs (33/53, 62%) used a statistical technique that took into account the repeated measurements of the data (e.g., time to event, linear mixed models) or statistically corrected for them if these repeated measures were tested independently from one another. Combined, these findings show that 41 of the 66 trials (41/66, 62%) addressed the issue of multiple testing either by statistically correcting for multiple domains, using a statistical technique that took into account the repeated measurements, or by analyzing only one follow-up time point. These findings remain consistent over time (2001-2006: 13/20, 65%; 2007-2012: 14/24, 58%; 2013-2017: 14/22, 64%).

The majority of the RCTs reported PRO scores descriptively (55/66, 83%), such as mean scores or mean change scores by trial arms, either on their own or as a support for a comparative analysis; and this has been quite consistent over the years (2001-2006: 16/20, 80%; 2007-2012: 19/24, 79%; 2013-2017: 20/22, 91%).

When analyzing PRO data, we identified more than six primary statistical analysis techniques. The top two most commonly used statistical techniques were (generalized) linear mixed models (18/66, 25%) and Wilcoxon ranks sums test/t-test (11/66, 17%). Many RCTs did not report the statistical technique used; a p-value was reported but it was not mentioned how this value was obtained (15/66, 23%). When comparing findings over time, the most commonly used statistical techniques between 2001-2006 were (generalized) linear mixed models (8/20, 40%) and Wilcoxon ranks sums test/t-test (5/20, 25%); between 2007-2012 were ANOVA/linear

regression (7/24, 29%), (generalized) linear mixed models (3/24, 13%) and Wilcoxon ranks sums test/t-test (3/24, 13%); and between 2013-2017 were (generalized) linear mixed models (7/22, 32%) and time to event (5/22, 23%). No single technique was used in a majority of the trials. Moreover, across all periods, a substantial proportion of RCTs failed to report the statistical technique used (2001-2006: 5/20, 25%; 2007-2012: 6/24, 25%; 2013-2017: 4/22, 18%).

Less than half of the RCTs addressed the clinical relevance of the findings (28/66, 42%). Among the trials that reported whether a finding was clinically relevant, the methods used varied: they were reported either as a change of X points from baseline (18/28, 64%), an X points difference between treatment arms (9/28, 32%) or both (1/28, 4%). The percentage of RCTs reporting the clinical relevance of their findings increased somewhat over the years (2001-2006: 5/20, 25%; 2007-2012: 11/24, 46%; 2013-2017: 12/22, 55%)

Baseline assessment

The majority of the RCTs included a baseline PRO assessment (60/66, 91%). From these 60 studies, 36 (36/60, 60%) compared PRO baseline scores between treatment arms and 13 (13/60, 22%) included the baseline score as a covariate. That the majority of the RCTs included a baseline PRO assessment has been consistent over the years (2001-2006: 18/20, 90%; 2007-2012: 22/24, 92%; 2013-2017: 20/22, 91%); however, the number of studies reporting whether PRO baseline scores are comparable between treatment arms seem to have declined over the years (2001-2006: 13/18, 72%; 2007-2012: 14/22, 64%; 2013-2017: 9/20, 45%); and including baseline scores as a covariate has not necessarily improved over the years (2001-2006: 2/18, 11%; 2007-2012: 6/22, 27%; 2013-2017: 5/20, 25%).

Amount of and handling of missing data

Many studies (24/66, 36%) did not report or did not clearly specify the analysis population for the primary PRO analysis; and this is still the case in the recent years (2001-2006: 6/20, 15%; 2007-2012: 8/24, 33%; 2013-2017: 10/22, 45%). Fourteen RCTs (14/66, 21%) reported using the intent-to-treat (ITT) population in their analysis; and a greater number of RCTs reported using a modified intent-to-treat (mITT) population (28/66, 42%). These numbers were relatively comparable over the years (see Table 2). Five different definitions of mITT were found, demonstrating that there is no consistent definition of mITT (64% with baseline PRO and ≥ 1 post-assessment (18/28); 14% with baseline PRO (4/28); 7% with at least one PRO data point (2/28); and 7% with baseline PRO and trial-specific follow-up point of interest (2/28). See Appendix, page 21-28, for the analysis population used by each RCT).

Regarding compliance rates, among the RCTs that assessed baseline PRO (60/66, 91%), twenty-eight of them (28/60, 47%) reported baseline PRO compliance rates for each treatment arm. Nineteen RCTs (19/66, 29%) reported whether compliance rates between treatment groups differed throughout the follow-up assessments. Most studies (48/66, 73%) did not report how missing data were dealt with. These findings were relatively comparable across the years (see Table 2).

Discussion

The aim of this systematic review was to assess the current state of PRO analysis in RCTs in advanced breast cancer. Our findings showed that in the 66 eligible RCTs, there was clear heterogeneity on how PRO data were analyzed.

Most trials failed to report a specific research hypothesis (88%), even in the last six years (2012-2017: 82%). This is consistent with previous reviews[18–21]. This may reflect lack of knowledge about the likely HRQOL trajectory for novel treatments or a lack of consideration of PRO specific hypotheses at the design stage and specification in the trial protocol. This is consistent with recent reviews of trial protocol content [22,23]. Our findings highlight an area of poor practice which does not meet ISOQOL and CONSORT-PRO reporting standards [24,25]. Failure to state a clear PRO hypothesis *a priori* opens up the possibility that inappropriate statistical techniques may be used. For instance, if a study had the objective about HRQOL changes over a six-week period, a cross-sectional HRQOL analysis at six weeks is not equivalent to an area under the curve analysis within the same time frame; in fact, it is possible that these two analytical techniques may yield different results. If the PRO objective is not stated or too vaguely stated, different statistical approaches may be reported as equivalent ways of addressing the same PRO objective, when in fact, they focus on different aspects of the data; and therefore respond to different research objectives. Divergent findings, however, may not necessarily invalidate the PRO data analysis but rather illustrate the importance of a well-defined *a priori* hypothesis, and responding to them with an appropriate statistical technique. Therefore, it is critical that researchers clearly define their hypotheses and appropriate corresponding statistical analyses in the protocol or statistical analysis plan in sufficient detail [26]; and results are described in a way

that accurately represents the key patterns in the data and able to be understood by non-statistical readers.

The most commonly used statistical technique (linear mixed models) was only employed in 27% of the RCTs (18/66). Wilcoxon-ranks-test/t-tests, statistical techniques appropriate for single time points or change scores, were also commonly used (11/66, 17%) although this strategy may not be appropriate since the majority of the trials involved analyzing data with more than two repeated assessments (53/66, 80%). There seems to be an increased interest in the use of time to event analysis in the recent years (from 2001-2007: 1/20, 5% to 2013-2017: 5/22, 23%) (see Table 2). However, a major concern remains that a number of RCTs (15/66, 23%) did not even (clearly) report the statistical technique they used to analyze PRO data, which is still evident in the recent years (2013-2017: 4/22, 18%).

Analysis of a PRO endpoint, such as HRQOL, often involves multiple outcomes. When drawing conclusions about treatment efficacy, it is advisable to avoid the risk of accumulating type 1 errors (false positive findings) by adjusting critical p-values for multiple comparisons when multiple outcomes are used to test a multi-dimensional endpoint, such as HRQOL. A large number of RCTs did not do this (30/38, 79%); and this has still been the case in the last six years (10/11, 91%), which may have led to erroneous conclusions about the PRO endpoint due to excess type 1 errors[27]. Given that results of these RCTs can lead to setting new standards of care, this practice should be avoided. On-going work from SPIRIT-PRO to standardize what needs to be included in the design stage of a trial (protocol) and statistical analysis plans may help promote better reporting on these issues [26].

The sample size estimation required for a trial is typically calculated only for the primary clinical endpoint. Since PRO endpoints, such as HRQOL, are often secondary endpoints, the sample size may be much larger (or smaller) than what is needed for that endpoint. Since statistical significance is highly dependent on sample size, having a large sample size can produce statistically significant results, but the clinical relevance of the change in the PRO endpoint may be negligible[28]. It is therefore recommended that clinical relevance should always be reported alongside statistical significance. Similar to other reviews [18–21,29], our review showed it is still not common practice to report the clinical relevance of PRO findings: less than half of the RCTs (28/66, 42%) reported whether their findings were clinically relevant; although this practice has shown some improvement in the last six years (from 2001-2006: 5/20, 25% to 2013-2017: 12/22, 55%).

The majority of the RCTs in this review reported having a baseline assessment (90%) and this has been consistent over the years. These findings demonstrate wide acceptance of this practice. Assessing baseline (or pre-treatment) scores is essential in any PRO analysis. Since individuals can differ in their baseline levels, it is important to take this into account when assessing individual changes over time and differences between treatment arms. This makes the statistical analysis more efficient by reducing the influence of baseline differences in the analysis[30]. A large number of articles collected baseline PRO information (60/66, 91%) and 40% of RCTs did not subsequently check whether there were baseline differences between treatment arms (24/60). Additionally, only a small number of trials reported using the baseline PRO scores as a covariate (13/60, 22%). These findings remain comparable over the years. This highlights the lack of consistency between investigators on how to use baseline information in their analyses.

To assess the amount of missing data, it is critical that trials report the set or subset of trial participants that will be used in the analysis (the "analysis population") [31], as well as PRO completion (or "compliance rates") over time[32]. Only a small number of the publications used intent-to-treat (ITT) as the analysis population (14/66, 21%); and this has still been the case in the recent years (2013-2017: 4/22, 18%). Additionally, some papers that purported to use ITT apparently did not adhere to the ITT principle (i.e., all randomized subjects should be analyzed according to the allocated treatment[33]). For example, some RCTs reported that they would use ITT for analysis, but their statistical techniques removed a patient if an assessment was missing (e.g., when a statistical test involves calculating a change score[34,35]). Probably because of the difficulty of using the ITT population for PRO analysis, a number of articles opted for a modified intent-to-treat approach (mITT). However, there is no consensus on which mITT approach should be used as demonstrated by the variety of ways these RCTs have defined their mITT (e.g., patients with baseline PRO; patients with baseline PRO + 1 follow-up assessment).

Compliance rates are another way of understanding the amount of missing data in a trial[32]. However, our findings showed that although more than half of the RCTs reported baseline compliance rates, a smaller number of publications reported follow-up compliance rates within their time frame of interest; and not all articles compared compliance rates between treatment groups. This lack of information on compliance rates makes it difficult to evaluate whether a statistical technique is appropriate for the analysis population (e.g., some statistical techniques assume that the dataset has no missing data or that missing data is missing completely at random) and whether the conclusions are generalizable to the population of interest.

Strategies to deal with missing data in the statistical analyses were reported in only 27% of RCTs (18/66); and this practice has not changed in the recent years (from 2001-2006: 4/20, 20% to 2013-2017: 5/22, 23%). However, it is known that missing data is a challenge in the analysis of PRO data in cancer trials[8,30,36]. As cancer patients often experience disease- and treatment-related illness and mortality, missing assessments are often inevitable[37]. Since missing data can bias results, it is strongly advised that sensitivity analyses should be conducted to explore the robustness of the primary findings [38]. That is, investigators are encouraged to reanalyze the data with a statistical model that makes different missing data assumptions than that of the primary analysis. If results are reasonably consistent across the different analyses, there is increased confidence that the presence of missing data did not compromise the original findings.[39] The lack of information on how missing data were handled suggests that this problem is often ignored or regarded as unimportant when reporting PRO findings. This situation should not be acceptable.

While our review was robust and followed a systematic approach, our work also has several limitations. Findings from this review were based on published articles, and the articles selected may reflect publication bias, i.e., statistically significant "positive" results tend to have a better chance of being published[40]. Protocols or *a priori* statistical analysis plans were not checked alongside these published reports. It is possible that information classified as "not reported" in this review may have been recorded in the protocol, but was not included in the article due to space limitations in the journals. However our findings are consistent with systematic reviews of protocols [22,23] and other reviews of papers reporting RCTs [18–21,29] demonstrating that these issues are indeed prevalent in the PRO field . We excluded non-English publications in our search, so some relevant trials may have been excluded. The focus of this systematic review was on

advanced breast cancer and thus may not be generalizable to all cancer types, although we have no reason to think that the analysis problems reported here would be different in other disease sites. Indeed, the converging results from other systematic reviews in different cancer sites point toward a general problem that is not specific to one cancer site[16,17,19]. As there are no agreed-upon standards on how to conduct analyses of PROs in RCTs, the evaluation criteria of these trials were based on authors' selection of statistical issues that were deemed as critical for the analysis of PRO data, but remains broadly in line with on-going work on guidelines for statistical analysis plans [26]. Although this review focuses on standards in statistical analysis, we would like to stress the importance of a high quality study design; and choosing appropriate PRO measures and assessment points that capture the impact of both the disease and treatment on the patient experience. Even if the most robust statistical approach is used, findings from a RCT would be of little relevance if the study design is of poor quality; and inappropriate outcomes and follow-up assessment points are used[26].

In conclusion, our review highlights the many statistical issues that need to be addressed to improve the analysis and interpretation of PRO data, including HRQOL. The lack of consensus on how to analyze PRO data makes it difficult to draw robust conclusions regarding PRO endpoints and compare findings across trials. Although the increased inclusion of PRO endpoints in RCTs is a substantial step toward a more patient-centered approach, standards and guidelines are needed for how to analyze PRO data in cancer RCTs. The SISAQOL Consortium was set up to address this need and develop recommendations on how to analyze PRO data in RCTs[11] and will produce such guidelines in the future.

References

1. Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington (DC): National Academy Press; 2001. doi:10.1200/JCO.2003.01.044.

2. Vodicka E, Kim K, Devine EB, Gnanasakthy A, Scoggins JF, Patrick DL. Inclusion of patient-reported outcome measures in registered clinical trials: Evidence from ClinicalTrials.gov (2007-2013). *Contemp Clin Trials*. 2015;43:1-9. doi:10.1016/j.cct.2015.04.004.

3. European Medicines Agency. Appendix 2. Guideline on the evaluation of anticancer medicinal products in man The use of patient-reported outcome (PRO) measures in oncology studies. http://www.ema.europa.eu/docs/en_GB/document_library/Other/2016/04/WC500205159.pdf. Published 2016. Accessed November 21, 2017.

4. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools). Silver Spring (MD): Food and Drug Administration (US); Bethesda (MD): National Institutes of Health (US). https://www.ncbi.nlm.nih.gov/books/NBK338448/. Published 2016. Accessed March 15, 2018.

5. Bottomley A. The Cancer Patient and Quality of Life. *Oncologist*. 2002;7(2):120-125. doi:10.1634/theoncologist.7-2-120.

6. LeBlanc TW, Abernethy AP. Patient-reported outcomes in cancer care — hearing the patient voice at greater volume. *Nat Rev Clin Oncol*. 2017;14(12):763-772. doi:10.1038/nrclinonc.2017.153.

7. Field KM, Jordan JT, Wen PY, Rosenthal MA, Reardon DA. Bevacizumab and glioblastoma: Scientific review, newly reported updates, and ongoing controversies. *Cancer*. 2015;121(7):997-1007. doi:10.1002/cncr.28935.

8. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient reported outcomes. *Stat Methods Med Res*. 2014;23(5):440-459. doi:10.1177/0962280213476378.

9. Fayers PM, Machin D. *Quality of Life: Assessment, Analysis and Interpretation of Patient-Reported Outcomes, 2nd Edtion.* Somerset, NJ: John Wiley & Sons; 2013.

10. Food and Drug Administration. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. https://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf. Published 2009. Accessed November 21, 2017.

11. Bottomley A, Pe M, Sloan J, et al. Analysing data from patient-reported outcome and quality of life endpoints for cancer clinical trials: a start in setting international standards. *Lancet Oncol*. 2016;17(11):e510-e514. doi:10.1016/S1470-2045(16)30510-1.

12. Cardoso F, Costa A, Norton L, et al. ESO-ESMO 2nd international consensus guidelines for advanced breast cancer (ABC2). *Ann Oncol*. 2014;25(10):1871-1888. doi:10.1093/annonc/mdu385.

13. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [Updated March 2011]*.; 2011. www.handbook.cochrane.org.

14. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group TP. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*.

2009;6(7):e1000097. doi:10.1371/journal.pmed.1000097.

15. Ghislain I, Zikos E, Coens C, et al. Health-related quality of life in locally advanced and metastatic breast cancer: methodological and clinical issues in randomised controlled trials. *Lancet Oncol*. 2016;17(7):e294-e304. doi:10.1016/S1470-2045(16)30099-7.

16. Fiteni F, Anota A, Westeel V, Bonnetain F. Methodology of health-related quality of life analysis in phase III advanced non-small-cell lung cancer clinical trials: a critical review. *BMC Cancer*. 2016;16:122. doi:10.1186/s12885-016-2152-1.

17. Hamel J-F, Saulnier P, Pe M, et al. A systematic review of the quality of statistical methods employed for analysing quality of life data in cancer randomised controlled trials. *Eur J Cancer*. 2017;83:166-176. doi:10.1016/j.ejca.2017.06.025.

18. Brundage M, Bass B, Davidson J, et al. Patterns of reporting health-related quality of life outcomes in randomized clinical trials: implications for clinicians and quality of life researchers. *Qual Life Res*. 2011;20(5):653-664. doi:10.1007/s11136-010-9793-3.

19. Mercieca-Bebber RL, Perreca A, King M, et al. Patient-reported outcomes in head and neck and thyroid cancer randomised controlled trials: A systematic review of completeness of reporting and impact on interpretation. *Eur J Cancer*. 2016;56:144-161. doi:10.1016/J.EJCA.2015.12.025.

20. Schandelmaier S, Conen K, von Elm E, et al. Planning and reporting of quality-of-life outcomes in cancer trials. *Ann Oncol*. 2015;26(9):1966-1973. doi:10.1093/annonc/mdv283.

21. Efficace F, Fayers P, Pusic A, et al. Quality of patient-reported outcome reporting across cancer randomized controlled trials according to the CONSORT patient-reported outcome

extension: A pooled analysis of 557 trials. *Cancer*. 2015;121(18):3335-3342. doi:10.1002/cncr.29489.

22. Kyte D, Duffy H, Fletcher B, et al. Systematic Evaluation of the Patient-Reported Outcome (PRO) Content of Clinical Trial Protocols. Briel M, ed. *PLoS One*. 2014;9(10):e110229. doi:10.1371/journal.pone.0110229.

23. Mercieca-Bebber R, Friedlander M, Kok P-S, et al. The patient-reported outcome content of international ovarian cancer randomised controlled trial protocols. *Qual Life Res*. 2016;25(10):2457-2465. doi:10.1007/s11136-016-1339-x.

24. Calvert M, Blazeby J, Altman DG, et al. Reporting of Patient-Reported Outcomes in Randomized Trials. *JAMA*. 2013;309(8):814. doi:10.1001/jama.2013.879.

25. Brundage M, Blazeby J, Revicki D, et al. Patient-reported outcomes in randomized clinical trials: Development of ISOQOL reporting standards. *Qual Life Res*. 2013;22(6):1161-1175. doi:10.1007/s11136-012-0252-1.

26. Calvert M, Kyte D, Mercieca-Bebber R, et al. Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension. *JAMA*. 2018;319(5):483. doi:10.1001/jama.2017.21903.

27. The European Agency for the Evaluation of Medicinal Products (EMEA). Points To Consider on Multiplicity Issues in Clinical Trials. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003640.pdf. Published 2002. Accessed November 21, 2017.

28. Ranganathan P, Pramesh CS, Buyse M. Common pitfalls in statistical analysis: Clinical versus statistical significance. *Perspect Clin Res*. 2015;6(3):169-170. doi:10.4103/2229-

3485.159943.

29. Cocks K, King MT, Velikova G, Fayers PM, Brown JM. Quality, interpretation and presentation of European Organisation for Research and Treatment of Cancer quality of life questionnaire core 30 data in randomised controlled trials. *Eur J Cancer*. 2008;44(13):1793-1798. doi:10.1016/J.EJCA.2008.05.008.

30. Fairclough DL. *Design and Analysis of Quality of Life Studies in Clinical Trials*. Florida: Chapman & Hall/CRC; 2002.

31. Chan A-W, Tetzlaff JM, Gotzsche PC, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ*. 2013;346(jan 08):e7586. doi:10.1136/bmj.e7586.

32. Mercieca-Bebber R, Palmer MJ, Brundage M, Calvert M, Stockler MR, King MT. Design, implementation and reporting strategies to reduce the instance and impact of missing patient-reported outcome (PRO) data: a systematic review. *BMJ Open*. 2016;6(6):e010938. doi:10.1136/bmjopen-2015-010938.

33. Montedori A, Bonacini MI, Casazza G, et al. Modified versus standard intention-to-treat reporting: Are there differences in methodological quality, sponsorship, and findings in randomized trials? A cross-sectional study. *Trials*. 2011;12(1):58. doi:10.1186/1745-6215-12-58.

34. Cassier PA, Chabaud S, Trillet-Lenoir V, et al. A phase-III trial of doxorubicin and docetaxel versus doxorubicin and paclitaxel in metastatic breast cancer: results of the ERASME 3 study. *Breast Cancer Res Treat*. 2007;109(2):343-350. doi:10.1007/s10549-007-9651-3.

35. Jones SE, Erban J, Overmoyer B, et al. Randomized phase III study of docetaxel compared with paclitaxel in metastatic breast cancer. *J Clin Oncol*. 2005;23(24):5542-5551. doi:10.1200/JCO.2005.02.027.

36. Rombach I, Rivero-Arias O, Gray AM, Jenkinson C, Burke Ó. The current practice of handling and reporting missing outcome data in eight widely used PROMs in RCT publications: a review of the current literature. *Qual Life Res*. 2016;25(7):1613-1623. doi:10.1007/s11136-015-1206-1.

37. Fairclough DL, Peterson HF, Chang V. Why are missing quality of life data a problem in clinical trials of cancer therapy? *Stat Med*. 1998;17(5-7):667-677.

38. Fielding S, Ogbuagu A, Sivasubramaniam S, MacLennan G, Ramsay CR. Reporting and dealing with missing quality of life data in RCTs: has the picture changed in the last decade? *Qual Life Res*. 2016;25(12):2977-2983. doi:10.1007/s11136-016-1411-6.

39. White IR, Horton NJ, Carpenter J, Pocock SJ, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. *BMJ*. 2011;342:d40. doi:10.1136/BMJ.D40.

40. Dubben H-H, Beck-Bornholdt H-P. Systematic review of publication bias in studies on publication bias. *BMJ*. 2005;331(7514):433-434. doi:10.1136/bmj.38478.497164.F7.

41. Burris HA, Lebrun F, Rugo HS, et al. Health-related quality of life of patients with advanced breast cancer treated with everolimus plus exemestane versus placebo plus exemestane in the phase 3, randomized, controlled, BOLERO-2 trial. *Cancer*. 2013;119(10):1908-1915. doi:10.1002/cncr.28010.

42. Campone M, Beck JT, Gnant M, et al. Health-related quality of life and disease symptoms

in postmenopausal women with HR(+), HER2(-) advanced breast cancer treated with everolimus plus exemestane versus exemestane monotherapy. *Curr Med Res Opin*. 2013;29(11):1463-1473. doi:10.1185/03007995.2013.836078.

43. Welslau M, Diéras V, Sohn JH, et al. Patient-reported outcomes from EMILIA, a randomized phase 3 study of trastuzumab emtansine (T-DM1) versus capecitabine and lapatinib in human epidermal growth factor receptor 2-positive locally advanced or metastatic breast cancer. *Cancer*. 2014;120(5):642-651. doi:10.1002/cncr.28465.

44. Hurvitz SA, Dirix L, Kocsis J, et al. Phase II randomized study of trastuzumab emtansine versus trastuzumab plus docetaxel in patients with human epidermal growth factor receptor 2-positive metastatic breast cancer. *J Clin Oncol*. 2013;31(9):1157-1163. doi:10.1200/JCO.2012.44.9694.

45. Svensson H, Einbeigi Z, Johansson H, Hatschek T, Brandberg Y. Quality of life in women with metastatic breast cancer during 9 months after randomization in the TEX trial (epirubicin and paclitaxel w/o capecitabine). *Breast Cancer Res Treat*. 2010;123(3):785-793. doi:10.1007/s10549-010-1084-8.

46. Brufsky A, Hoelzer K, Beck T, et al. A randomized phase II study of paclitaxel and bevacizumab with and without gemcitabine as first-line treatment for metastatic breast cancer. *Clin Breast Cancer*. 2011;11(4):211-220. doi:10.1016/j.clbc.2011.03.019.

47. Thomas ES, Gomez HL, Li RK, et al. Ixabepilone plus capecitabine for metastatic breast cancer progressing after anthracycline and taxane treatment. *J Clin Oncol*. 2007;25(33):5210-5217. doi:10.1200/JCO.2007.12.6557.

48. Corey-Lisle PK, Peck R, Mukhopadhyay P, et al. Q-TWiST analysis of ixabepilone in

combination with capecitabine on quality of life in patients with metastatic breast cancer. *Cancer*. 2012;118(2):461-468. doi:10.1002/cncr.26213.

49. Nuzzo F, Morabito A, Gravina A, et al. Effects on quality of life of weekly docetaxel-based chemotherapy in patients with locally advanced or metastatic breast cancer: results of a single-centre randomized phase 3 trial. *BMC Cancer*. 2011;11(1):75. doi:10.1186/1471-2407-11-75.

50. Di Leo A, Jerusalem G, Petruzelka L, et al. Results of the CONFIRM phase III trial comparing fulvestrant 250 mg with fulvestrant 500 mg in postmenopausal women with estrogen receptor-positive advanced breast cancer. *J Clin Oncol*. 2010;28(30):4594-4600. doi:10.1200/JCO.2010.28.8415.

51. Ellis MJ, Gao F, Dehdashti F, et al. Lower-dose (6 mg Daily) versus High-dose (30 mg Daily) Oral Estradiol Therapy of Hormone-receptor-positive, Aromatase- inhibitor-resistant Advanced Breast Cancer: A Randomized Phase 2 Study. *JAMA*. 2009;302(7):774-780. doi:10.1001/jama.2009.1204.Lower-dose.

52. Zhou X, Cella D, Cameron D, et al. Lapatinib plus capecitabine versus capecitabine alone for HER2+ (ErbB2+) metastatic breast cancer: Quality-of-life assessment. *Breast Cancer Res Treat*. 2009;117(3):577-589. doi:10.1007/s10549-009-0310-8.

53. Hopwood P, Watkins J, Ellis P, Smith I. Clinical interpretation of quality-of-life outcomes: An investigation of data from the randomized trial of gemcitabine plus paclitaxel compared with paclitaxel alone for advanced breast cancer. *Breast J*. 2008;14(3):228-235. doi:10.1111/j.1524-4741.2008.00567.x.

54. Moinpour CM, Donaldson GW, Liepa AM, Melemed AS, O'Shaughnessy J, Albain KS.

Evaluating health-related quality-of-life therapeutic effectiveness in a clinical trial with extensive nonignorable missing data and heterogeneous response: results from a phase III randomized trial of gemcitabine plus paclitaxel versus paclitaxel monothe. *Qual Life Res*. 2012;21(5):765-775. doi:10.1007/s11136-011-9999-z.

55. Chia S, Gradishar W, Mauriac L, et al. Double-Blind, Randomized Placebo Controlled Trial of Fulvestrant Compared With Exemestane After Prior Nonsteroidal Aromatase Inhibitor Therapy in Postmenopausal Women With Hormone Receptor–Positive, Advanced Breast Cancer: Results From EFECT. *J Clin Oncol*. 2008;26(10):1664-1670. doi:10.1200/JCO.2007.13.5822.

56. Reyno L. Phase III Study of N , N -Diethyl-2- [4- (Phenylmethyl) Phenoxy] Ethanamine (BMS-217380-01) Combined With Doxorubicin Versus Doxorubicin Alone in Metastatic / Recurrent Breast Cancer : National Cancer Institute of Canada Clinical Trials Group Study. *J Clin Oncol*. 2004;22(2):269-276. doi:10.1200/JCO.2004.04.075.

57. Liu J, Tu D, Dancey J, et al. Quality of life analyses in a clinical trial of DPPE (tesmilifene) plus doxorubicin versus doxorubicin in patients with advanced or metastatic breast cancer: NCIC CTG Trial MA.19. *Breast Cancer Res Treat*. 2006;100(3):263-271. doi:10.1007/s10549-006-9257-1.

58. Therasse P, Mauriac L, Welnicka-Jaskiewicz M, et al. Final Results of a Randomized Phase III Trial Comparing Cyclophosphamide, Epirubicin, and Fluorouracil With a Dose-Intensified Epirubicin and Cyclophosphamide + Filgrastim as Neoadjuvant Treatment in Locally Advanced Breast Cancer: An EORTC-NCIC-SAKK Mult. *J Clin Oncol*. 2003;21(5):843-850. doi:10.1200/JCO.2003.05.135.

59. Bottomley A, Therasse P, Piccart M, et al. Health-related quality of life in survivors of locally advanced breast cancer: An international randomised controlled phase III trial. *Lancet Oncol*. 2005;6(5):287-294. doi:10.1016/S1470-2045(05)70100-5.

60. Fountzilas G, Kalofonos HP, Dafni U, et al. Paclitaxel and epirubicin versus paclitaxel and carboplatin as first-line chemotherapy in patients with advanced breast cancer: A phase III study conducted by the Hellenic Cooperative Oncology Group. *Ann Oncol*. 2004;15(10):1517-1526. doi:10.1093/annonc/mdh395.

61. Howell A, Robertson JFR, Albano JQ, et al. Fulvestrant, formerly ICI 182,780, is as effective as anastrozole in postmenopausal women with advanced breast cancer progressing after prior endocrine treatment. *J Clin Oncol*. 2002;20(16):3396-3403. doi:10.1200/JCO.2002.10.057.

62. Osborne CK, Pippen J, Jones SE, et al. Double-blind, randomized trial comparing the efficacy and tolerability of fulvestrant versus anastrozole in postmenopausal women with advanced breast cancer progressing on prior endocrine therapy: Results of a North American trial. *J Clin Oncol*. 2002;20(16):3386-3395. doi:10.1200/JCO.2002.10.058.

63. Buzdar A, Douma J, Davidson N, et al. Phase III, multicenter, double-blind, randomized study of letrozole, an aromatase inhibitor, for advanced breast cancer versus megestrol acetate. *J Clin Oncol*. 2001;19(14):3357-3366. doi:10.1200/JCO.2001.19.14.3357.

64. Conte PF, Guarneri V, Bruzzi P, et al. Concomitant versus sequential administration of epirubicin and paclitaxel as first-line therapy in metastatic breast carcinoma: Results from the gruppo oncologico nord ovest randomized trial. *Cancer*. 2004;101(4):704-712. doi:10.1002/cncr.20400.

65. Keller AM, Mennel RG, Georgoulias VA, et al. Randomized phase III trial of pegylated liposomal doxorubicin versus vinorelbine or mitomycin C plus vinblastine in women with taxane-refractory advanced breast cancer. *J Clin Oncol*. 2004;22(19):3893-3901. doi:10.1200/JCO.2004.08.157.

66. O'Shaughnessy J, Miles D, Vukelja S, et al. Superior survival with capecitabine plus docetaxel combination therapy in anthracycline-pretreated patients with advanced breast cancer: Phase III trial results. *J Clin Oncol*. 2002;20(12):2812-2823. doi:10.1200/JCO.2002.09.002.

67. Cortés J, Baselga J, Im Y-H, et al. Health-related quality-of-life assessment in CLEOPATRA, a phase III study combining pertuzumab with trastuzumab and docetaxel in metastatic breast cancer. *Ann Oncol*. 2013;24(10):2630-2635. doi:10.1093/annonc/mdt274.

68. Lück H-J, Du Bois A, Loibl S, et al. Capecitabine plus paclitaxel versus epirubicin plus paclitaxel as first-line treatment for metastatic breast cancer: efficacy and safety results of a randomized, phase III trial by the AGO Breast Cancer Study Group. *Breast Cancer Res Treat*. 2013;139(3):779-787. doi:10.1007/s10549-013-2589-8.

69. Gianni L, Romieu GH, Lichinitser M, et al. AVEREL: A randomized phase III trial evaluating bevacizumab in combination with docetaxel and trastuzumab as first-line therapy for her2-positive locally recurrent/metastatic breast cancer. *J Clin Oncol*. 2013;31(14):1719-1725. doi:10.1200/JCO.2012.44.7912.

70. Bachelot T, Bajard A, Ray-Coquard I, et al. Final results of ERASME-4: A randomized trial of first-line docetaxel plus either capecitabine or epirubicin for metastatic breast

cancer. *Oncology*. 2011;80(3-4):262-268. doi:10.1159/000329066.

71. Blackwell KL, Burstein HJ, Storniolo AM, et al. Randomized study of lapatinib alone or in combination with trastuzumab in women with ErbB2-positive, trastuzumab-refractory metastatic breast cancer. *J Clin Oncol*. 2010;28(7):1124-1130. doi:10.1200/JCO.2008.21.4437.

72. Wu Y, Amonkar MM, Sherrill BH, et al. Impact of lapatinib plus trastuzumab versus single-agent lapatinib on quality of life of patients with trastuzumab-refractory HER2+ metastatic breast cancer. *Ann Oncol*. 2011;22(12):2582-2590. doi:10.1093/annonc/mdr014.

73. Schröder CP, De Munck L, Westermann AM, et al. Weekly docetaxel in metastatic breast cancer patients: No superior benefits compared to three-weekly docetaxel. *Eur J Cancer*. 2011;47(9):1355-1362. doi:10.1016/j.ejca.2010.12.018.

74. Sherrill B, Amonkar MM, Sherif B, Maltzman J, O'Rourke L, Johnston S. Quality of Life in Hormone Receptor-Positive HER-2+ Metastatic Breast Cancer Patients During Treatment with Letrozole Alone or in Combination with Lapatinib. *Oncologist*. 2010;15(9):944-953. doi:10.1634/theoncologist.2010-0012.

75. Sherrill B, Di Leo A, Amonkar MM, et al. Quality-of-life and quality-adjusted survival (Q-TWiST) in patients receiving lapatinib in combination with paclitaxel as first-line treatment for metastatic breast cancer. *Curr Med Res Opin*. 2010;26(4):767-775. doi:10.1185/03007991003590860.

76. Meier CR, Illiger HJ, Steder M, et al. Weekly vinorelbine versus docetaxel for metastatic breast cancer after failing anthracycline treatment. *Onkologie*. 2008;31(8-9):447-453.

doi:10.1159/000140453.

77. Fountzilas G, Dafni U, Dimopoulos MA, et al. A randomized phase III study comparing three anthracycline-free taxane-based regimens, as first line chemotherapy, in metastatic breast cancer: A Hellenic Cooperative Oncology Group study. *Breast Cancer Res Treat.* 2009;115(1):87-99. doi:10.1007/s10549-008-0047-9.

78. Miller K, Wang M, Gralow J, et al. Paclitaxel plus Bevacizumab versus Paclitaxel Alone for Metastatic Breast Cancer. *N Engl J Med*. 2007;357(26):2666-2676. doi:10.1056/NEJMoa072113.

79. Cella D, Wang M, Wagner L, Miller K. Survival-adjusted health-related quality of life (HRQL) among patients with metastatic breast cancer receiving paclitaxel plus bevacizumab versus paclitaxel alone: results from Eastern Cooperative Oncology Group Study 2100 (E2100). *Breast Cancer Res Treat*. 2011;130(3):855-861. doi:10.1007/s10549-011-1725-6.

80. Crump M, Gluck S, Tu D, et al. Randomized trial of high-dose chemotherapy with autologous peripheral-blood stem-cell support compared with standard-dose chemotherapy in women with metastatic breast cancer: NCIC MA.16. *J Clin Oncol*. 2008;26(1):37-43. doi:10.1200/JCO.2007.11.8851.

81. Karamouzis M V., Ioannidis G, Rigatos G. Quality of life in metastatic breast cancer patients under chemotherapy or supportive care: A single-institution comparative study. *Eur J Cancer Care (Engl)*. 2007;16(5):433-438. doi:10.1111/j.1365-2354.2006.00771.x.

82. von Minckwitz G, Chernozemsky I, Sirakova L, et al. Bendamustine prolongs progression-free survival in metastatic breast cancer (MBC): A phase III prospective,

randomized, multicenter trial of bendamustine hydrochloride, methotrexate and 5-fluorouracil (BMF) versus cyclophosphamide, methotrexate and 5-fluo. *Anticancer Drugs*. 2005;16(8):871-877. doi:10.1097/01.cad.0000175587.31940.19.

83. Miller KD, Chap LI, Holmes FA, et al. Randomized phase III trial of capecitabine compared with bevacizumab plus capecitabine in patients with previously treated metastatic breast cancer. *J Clin Oncol*. 2005;23(4):792-799. doi:10.1200/JCO.2005.05.098.

84. Bottomley A, Biganzoli L, Cufer T, et al. Randomized , Controlled Trial Investigating Short-Term Health-Related Quality of Life With Doxorubicin and Paclitaxel Versus Doxorubicin and Cyclophosphamide As First-Line Chemotherapy in Patients With Metastatic Breast Cancer : European Organization for. *J Clin Oncol*. 2004;22(13):2576-2586. doi:10.1200/JCO.2004.02.037.

85. Winer EP, Berry DA, Woolf S, et al. Failure of higher-dose paclitaxel to improve outcome in patients with metastatic breast cancer: Cancer and leukemia group B trial 9342. *J Clin Oncol*. 2004;22(11):2061-2068. doi:10.1200/JCO.2004.08.048.

86. Nabholtz JM, Falkson C, Campos D, et al. Docetaxel and doxorubicin compared with doxorubicin and cyclophosphamide as first-line chemotherapy for metastatic breast cancer: Results of a randomized, multicenter, phase III trial. *J Clin Oncol*. 2003;21(6):968-975. doi:10.1200/JCO.2003.04.040.

87. Sledge GW, Neuberg D, Bernardo P, et al. Phase III Trial of Doxorubicin, Paclitaxel, and the Combination of Doxorubicin and Paclitaxel as Front-Line Chemotherapy for Metastatic Breast Cancer: An Intergroup Trial (E1193). *J Clin Oncol*. 2003;21(4):588-

592. doi:10.1200/JCO.2003.08.013.

88. Osoba D, Slamon DJ, Burchmore M, Murphy M. Effects on quality of life of combined trastuzumab and chemotherapy in women with metastatic breast cancer. *J Clin Oncol*. 2002;20(14):3106-3113. doi:10.1200/JCO.2002.03.090.

89. Vogel CL, Cobleigh MA, Tripathy D, et al. Efficacy and Safety of Trastuzumab as a Single Agent in First-Line Treatment of *HER2* -Overexpressing Metastatic Breast Cancer. *J Clin Oncol*. 2002;20(3):719-726. doi:10.1200/JCO.2002.20.3.719.

90. Rugo H, Brammer M, Zhang F, Lalla D. Effect of Trastuzumab on Health-Related Quality of Life in Patients With HER2-Positive Metastatic Breast Cancer: Data From Three Clinical Trials. *Clin Breast Cancer*. 2010;10:288-293. doi:10.3816/CBC.2010.n.037.

91. Eiermann W. Trastuzumab combined with chemotherapy for the treatment of. *Ann Oncol*. 2001;12(1):57-62.

92. Chan S, Romieu G, Huober J, et al. Phase III study of gemcitabine plus docetaxel compared with capecitabine plus docetaxel for anthracycline-pretreated patients with metastatic breast cancer. *J Clin Oncol*. 2009;27(11):1753-1760. doi:10.1200/JCO.2007.15.8485.

93. Del Mastro L, Fabi A, Mansutti M, et al. Randomised phase 3 open-label trial of first-line treatment with gemcitabine in association with docetaxel or paclitaxel in women with metastatic breast cancer: a comparison of different schedules and treatments. *BMC Cancer*. 2013;13(1):164. doi:10.1186/1471-2407-13-164.

94. Park YH, Jung KH, Im S -a., et al. Phase III, Multicenter, Randomized Trial of Maintenance Chemotherapy Versus Observation in Patients With Metastatic Breast

Cancer After Achieving Disease Control With Six Cycles of Gemcitabine Plus Paclitaxel As First-Line Chemotherapy: KCSG-BR07-02. *J Clin Oncol*. 2013;31(14):1732-1740. doi:10.1200/JCO.2012.45.2490.

95. Park YH, Jung KH, Im SA, et al. Quality of life (QoL) in metastatic breast cancer patients with maintenance paclitaxel plus gemcitabine (PG) chemotherapy: results from phase III, multicenter, randomized trial of maintenance chemotherapy versus observation (KCSG-BR07-02). *Breast Cancer Res Treat*. 2015;(152):77-85. doi:10.1200/JCO.2012.45.2490.

96. Gelmon KA, Boyle FM, Kaufman B, et al. Lapatinib or trastuzumab plus taxane therapy for human epidermal growth factor receptor 2-positive advanced breast cancer: Final results of NCIC CTG MA.31. *J Clin Oncol*. 2015;33(14):1574-1583. doi:10.1200/JCO.2014.56.9590.

97. Kaufman PA, Awada A, Twelves C, et al. Phase III open-label randomized study of eribulin mesylate versus capecitabine in patients with locally advanced or metastatic breast cancer previously treated with an anthracycline and a taxane. *J Clin Oncol*. 2015;33(6):594-601. doi:10.1200/JCO.2013.52.4892.

98. Cortes J, Hudgens S, Twelves C, et al. Health-related quality of life in patients with locally advanced or metastatic breast cancer treated with eribulin mesylate or capecitabine in an open-label randomized phase 3 trial. *Breast Cancer Res Treat*. 2015;154(3):509-520. doi:10.1007/s10549-015-3633-7.

99. Park IH, Ro J, Lee KS, Kim SN, Yun YH, Nam BH. Phase II study of gemcitabine in combination with vinorelbine versus gemcitabine followed by vinorelbine for metastatic breast cancer. *Invest New Drugs*. 2010;28(5):659-669. doi:10.1007/s10637-009-9285-x.

100. Pivot X, Spano JP, Espie M, et al. Patients' preference of trastuzumab administration (subcutaneous versus intravenous) in HER2-positive metastatic breast cancer: Results of the randomised MetaspHer study. *Eur J Cancer*. 2017;82:230-236. doi:10.1016/j.ejca.2017.05.009.

101. Perez EA, Barrios C, Eiermann W, et al. Trastuzumab emtansine with or without pertuzumab versus trastuzumab plus taxane for human epidermal growth factor receptor 2-positive, advanced breast cancer: Primary results from the phase III MARIANNE study. *J Clin Oncol*. 2017;35(2):141-148. doi:10.1200/JCO.2016.67.4887.

102. Shiroiwa T, Fukuda T, Shimozuma K, et al. Long-term health status as measured by EQ-5D among patients with metastatic breast cancer: comparison of first-line oral S-1 and taxane therapies in the randomized phase III SELECT BC trial. *Qual Life Res*. 2017;26(2):445-453. doi:10.1007/s11136-016-1388-1.

103. Cinieri S, Chan A, Altundag K, et al. Final Results of the Randomized Phase II NorCap-CA223 Trial Comparing First-Line All-Oral Versus Taxane-Based Chemotherapy for HER2-Negative Metastatic Breast Cancer. *Clin Breast Cancer*. 2017;17(2):91-99. doi:10.1016/j.clbc.2016.06.014.

104. Rochlitz C, Bigler M, von Moos R, et al. SAKK 24/09: safety and tolerability of bevacizumab plus paclitaxel vs. bevacizumab plus metronomic cyclophosphamide and capecitabine as first-line therapy in patients with HER2- negative advanced stage breast cancer - a multicenter, randomized phase III trial. *BMC Cancer*. 2016;16. doi:10.1186/s12885-016-2823-y.

105. Gradishar WJ, Tjulandin S, Davidson N, et al. Phase III trial of nanoparticle albumin-

bound paclitaxel compared with polyethylated castor oil-based paclitaxel in women with breast cancer. *J Clin Oncol*. 2005;23(31):7794-7803. doi:10.1200/JCO.2005.04.937.

106. Howell A, Robertson JFR, Abram P, et al. Comparison of fulvestrant versus tamoxifen for the treatment of advanced breast cancer in postmenopausal women previously untreated with endocrine therapy: a multinational, double-blind, randomized trial. *J Clin Oncol*. 2004;22(9):1605-1613. doi:10.1200/JCO.2004.02.112.

107. Takashima T, Mukai H, Hara F, et al. Taxanes versus S-1 as the first-line chemotherapy for metastatic breast cancer (SELECT BC): an open-label, non-inferiority, randomised phase 3 trial. *Lancet Oncol*. 2016;17(1):90-98. doi:10.1016/S1470-2045(15)00411-8.

108. Yamamoto D, Sato N, Rai Y, et al. Efficacy and safety of low-dose capecitabine plus docetaxel versus single-agent docetaxel in patients with anthracycline-pretreated HER2-negative metastatic breast cancer: results from the randomized phase III JO21095 trial. *Breast Cancer Res Treat*. 2017;161(3):473-482. doi:10.1007/s10549-016-4075-6.

109. Pagani O, Klingbiel D, Ruhstaller T, et al. Do all patients with advanced HER2 positive breast cancer need upfront-chemo when receiving trastuzumab? Randomized phase III trial SAKK 22/99. *Ann Oncol*. 2017;28(2):305-312. doi:10.1093/annonc/mdw622.

110. Harbeck N, Saupe S, Jäger E, et al. A randomized phase III study evaluating pegylated liposomal doxorubicin versus capecitabine as first-line therapy for metastatic breast cancer: results of the PELICAN study. *Breast Cancer Res Treat*. 2017;161(1):63-72. doi:10.1007/s10549-016-4033-3.

111. Vrdoljak E, Marschner N, Zielinski C, et al. Final results of the TANIA randomised phase III trial of bevacizumab after progression on first-line bevacizumab therapy for HER2-

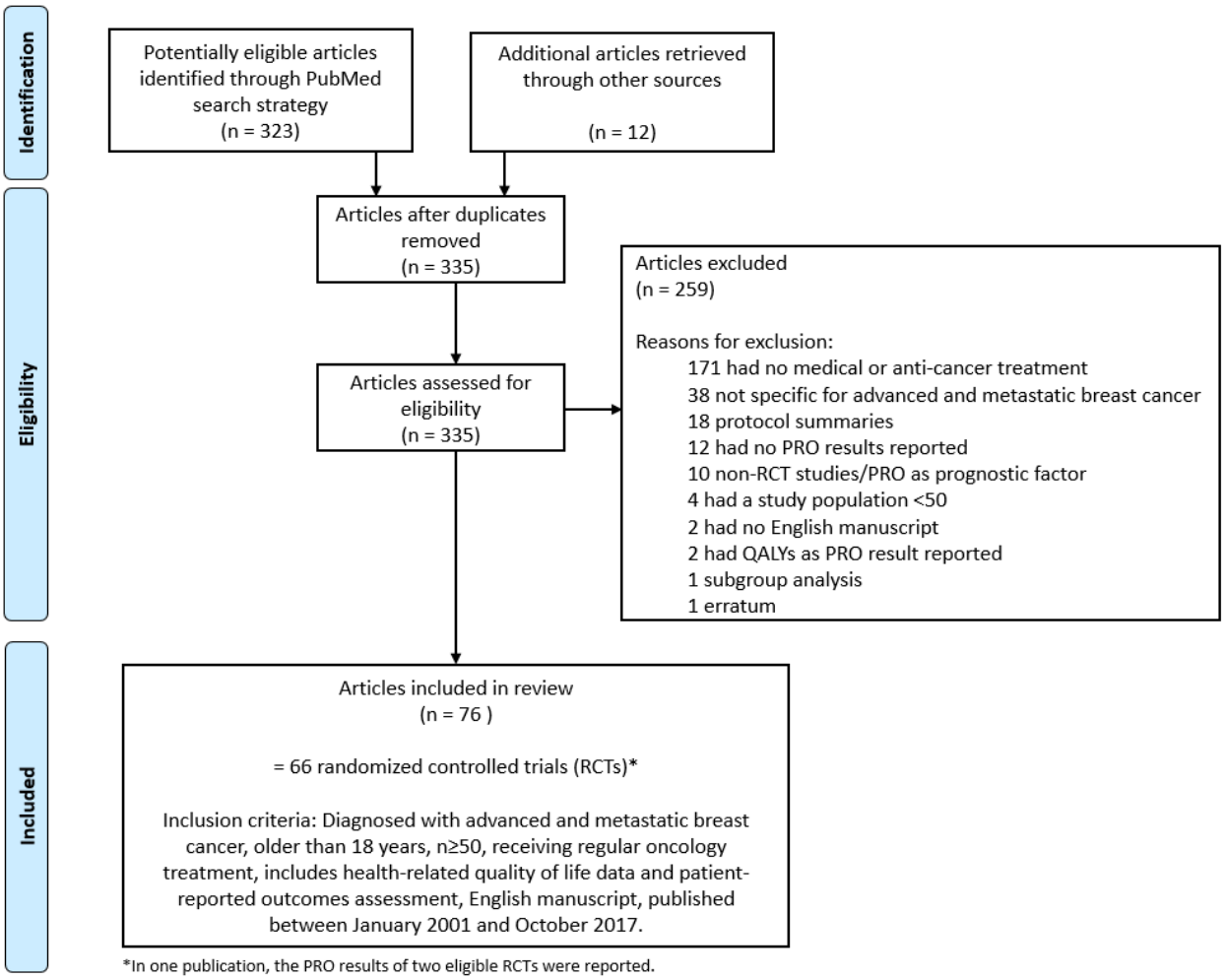
negative locally recurrent/metastatic breast cancer. *Ann Oncol*. 2016;27(11):2046-2052. doi:10.1093/annonc/mdw316.

112. Turner NC, Ro J, André F, et al. Palbociclib in Hormone-Receptor–Positive Advanced Breast Cancer. *N Engl J Med*. 2015;373(3):209-219. doi:10.1056/NEJMoa1505270.

113. Harbeck N, Iyer S, Turner N, et al. Quality of life with palbociclib plus fulvestrant in previously treated hormone receptor-positive, HER2-negative metastatic breast cancer: Patient-reported outcomes from the PALOMA-3 trial. *Ann Oncol*. 2016;27(6):1047-1054. doi:10.1093/annonc/mdw139.

114. Bell T, Crown JP, Lang I, et al. Impact of palbociclib plus letrozole on pain severity and pain interference with daily activities in patients with estrogen receptor-positive/human epidermal growth factor receptor 2-negative advanced breast cancer as first-line treatment. *Curr Med Res Opin*. 2016;32(5):959-965. doi:10.1185/03007995.2016.1157060.

Figure 1: **Search Strategy flowchart for the inclusion and exclusion of RCTs**



Authors' Contributions

All authors conceptualized the idea during the SISAQOL consortium meeting in Brussels in January 2016. M.Pe, and C. Coens conceptualized and developed the relevant statistical issues needed to be assessed for the analysis of PRO data. M.Pe carried out the systematic review with L. Dorme as the second reviewer. M.Pe, L. Dorme, C. Coens, A. Bottomley contributed to the initial interpretation of the results. M.Pe took the lead in drafting the manuscript. M. Pe and L. Dorme drafted the initial summary of findings. L. Dorme took the lead in the presentation of the raw results found in the Appendix. A. Bottomley supervised the findings and writing of this work. All authors discussed the results, provided critical feedback and reviewed the manuscript. All authors approved the final draft of the manuscript.

Conflict of Interest Statement

AB reports grants from Boehringer Ingelheim, grants from EORTC cancer research fund, during the conduct of the study; grants from Merck, outside the submitted work; and member of the EORTC Quality of Life Group executive committee. AC reports other from Genentech, A Member of the Roche Group, employee, outside the submitted work. GV reports personal fees and non-financial support from Roche, personal fees and non-financial support from Eisai, personal fees from Novartis, grants from National Institute Health Research England, grants from Yorkshire Cancer Research, grants from Breast Cancer Now, grants from EORTC Quality of Life Group, outside the submitted work. IG reports being an employee of Boehringer Ingelheim which provided an unrestricted education grant to EORTC. KO reports grants for the International Brain Tumour Alliance (IBTA) from AbbVie, Accuray, Antisense Pharma, Apogenix, Archimedes, Ark Therapeutics, Astra Zeneca, Boehringer Ingelheim, Brain Tumour Network (USA), Brain Tumor Resource and Information Network (USA), Bristol-Myers Squbb (BMS) Celldex Therapeutics, Crusade, Dijon Designs (UK), Elekta, Eli Lilly, Gerry & Nancy Pencer Brain Trust (Canada), Gosling Foundation (UK), GlaxoSmithKline (GSK), Ivy Foundation (USA), Lully, Link Pharmaceuticals, MagForce, Medac, Merck Serono, Merck, MGI Pharma, MSD Oncology, NeoPharm, Neuroendoscopy (Australia), Northwest Biotherapeutics, Novartis, Novocure, Pediatric Brain Tumor Foundation (USA), Pfizer, Photonamic, Roche, Schering-Plough (Global), Sontag Foundation (USA), Spink (UK), to-BBB, Vane Percy (UK), VBL Therapeutics and the Wallerstein Foundation (USA), all of which are outside the submitted work. KC reports other from Amgen, other from BMS, other from Celgene, other from Adelphi Values, other from Endomag, outside the submitted work. MC reports personal fees from Astellas, grants from NIHR, outside the submitted work; and International Society for Quality of Life Research, Best Practices for PRO in Trials Taskforce Chair. MKo reports grants from EORTC, Biofrontera, KFN, personal fees from Janssen-Cilag outside the submitted work. ND reports grants from the EuroQol Group, and grants from Association of the British Pharmaceutical Industry outside the submitted work

All other authors have no conflict of interest to disclose.

This study received no NIH funding. None of the authors was fully or partly NIH funded, employed by NIH or are receipt of an NIH grant.